# FREQUENTIST PRIORS

By Jiunn T. Hwang[1] and George Casella[2]

*Cornell University*

BU-924-M May, 1987

*AMS* 1980 *subject classifications.* 62F99, 62F25

*Key words and phrases.* Conditional inference, Bayes.

[1] Research supported by National Science Foundation Grant No. DMS8506847.

[2] Research supported by National Science Foundation Grant No. DMS8501973.

## SUMMARY

One way of attaining coherent post-data probability statements is through the use of posterior probabilities based on proper prior distributions. Classical frequency theory does not recognize the use of priors and, in some cases, coherent post-data inference cannot be made. Bayesian theory, however, does depend on the use of priors, and if one is to reconcile these two points of view in the set estimation problem, one needs a prior that is acceptable to a frequentist. Such a prior, which can lead to reconciliation, is called a frequentist prior, and can be used as a basis for coherent post-data inference. Existence and other properties of these priors are investigated.

**1. Introduction.** In the frequency theory of statistics, all probability statements are based on a repeated sampling framework. These probabilities are pre-data probabilities; in particular, confidence statements can be asserted before the data are seen. Furthermore, once the data have been seen the frequentist is left with nothing. In the words of LeCam (1984), frequency theory "··· does not have any probabilities to play with once the dice have been cast."

Once the dice have been cast, or once the data have been collected, the frequentist has nothing random left to base probability statements on. This is a shortcoming of frequency theory for at least two reasons. One, the more practical, is that users want to make post-data probability statements, and will do so even without a legitimate probability structure. Two, the more philosophical reason, is that frequentist probability assertions ignore the observed data, and surely one should be able to produce better inferences by not ignoring this information.

The problem lies not in the necessity of post-data inference: we believe that most statisticians would agree that post-data inference is a desirable thing. The problem lies in the mechanism available for frequentist post-data probability. As LeCam states, there is none.

Theories of conditional inference, which can provide a type of post-data assessment, have been around for a long time (Fisher, 1956; Buehler, 1959; Kiefer, 1977; Robinson, 1979). Most recently, Casella (1985,1986) has attempted to interpret the pre-data frequentist confidence in a post-data setting, using the theory of relevant betting procedures. These conditional theories attempt to give an objective post-data assessment, but there is an underlying connection in this work: coherent post-data assessments come from Bayes or limits of Bayes procedures, with the strongest conditional properties belonging to procedures based on proper priors (Robinson, 1979; Pierce, 1973).

Why is coherent post-data inference so intertwined with Bayes procedures? The mathematics is inarguable, but the intuition is even more illuminating: the only probability available post-data is subjective probability, so post-data inferences must necessarily be subjective. The frequentist is backed into a Bayesian corner: he must formulate a subjective opinion on which to base post-data statements.

The frequentist is, therefore, in the (somewhat uncomfortable) position of having to formulate a prior if he wants to make coherent post-data statements. Of course, noninformative or other improper priors can be used, but this tactic really avoids the primary issue. If one is forced to use subjective information, why not do the best possible job?

In this paper we investigate the types of proper priors that might be acceptable to a frequentist. We make some broad assumptions on the prior, and assume that a prior median can be specified. Other than that, the form of the prior is left unspecified. We also confine ourselves, for the most part, to the one parameter location case, although extensions to a case with nuisance parameters, or to higher dimensions, are possible.

Given that an experimenter observes X according to $f(x-\theta)$, and has a prior $\pi(\theta)$ with median $\mu$, we can loosely define three possible cases:

I. The data and the prior agree.

II. The data and the prior disagree moderately.

III. The data and the prior disagree greatly.

In Case I there is no problem: the sample density and the prior are quite close, and the subjective post-data assessment will agree with a pre-data assessment. In fact, the post-data confidence level will be greater than the pre-data one. Case II poses many problems, most of which will not be answered here. Pre-data and post-data confidence levels will be discrepant, and belief in the post-data probability necessitates belief in the form of the prior, a step the frequentist may not want to take.

Case III can be resolved, however, and its resolution provides us with a class of priors that may be acceptable to the frequentist. If the data and the prior disagree greatly, the frequentist has only one choice: believe the data. More precisely, if $C = \{\theta: |\theta-X| \leq c\}$ is a $1-\alpha$ frequentist confidence set, then the posterior probability of C must satisfy

$$\lim_{|X-\mu|\to\infty} P(\theta \in C|X) = 1-\alpha \quad , \tag{1.1}$$

where $\mu$ is the prior median. Note that condition (1.1) is different from a standard Bayesian asymptotic result: If $\bar{X}_n$ is the mean of a sample of size n, then

$$\lim_{n\to\infty} P(\theta \in C_n|\bar{X}_n) = 1-\alpha \quad , \tag{1.2}$$

where $C_n = \{\theta: |\theta-\bar{X}_n| \leq c/\sqrt{n}\}$. Condition (1.2) is essentially a variance condition: as the sampling variance $\to 0$ the prior washes out. Condition (1.1) compares two proper densities with finite variances (in most cases), and need not hold even if (1.2) does. If $f(x-\theta)$ is $n(\theta,1)$ and $\pi(\theta)$ is $n(0,1)$, then (1.2) but not (1.1) holds.

A problem similar to the one described here has previously been addressed by Dawid (1973), and later by Meeden and Isaacson (1977). Generalizing Dawid's work, Meeden and Isaacson are concerned with the "large X" behavior of a posterior distribution, and find conditions under which these "large X" posteriors are normal. They also study the behavior of the Bayes point estimate. In the present work our concern is with "large X" behavior of confidence intervals, not point estimates.

We consider a prior to be minimally acceptable to a frequentist if (1.1) holds, and thus it can be used as a basis for coherent post-data inference. This inference, used in conjunction with a pre-data confidence level, can provide a total inference for a frequentist, one that uses both the repeated sampling framework (pre-data) and a subjective framework (post-data).

**2. Preliminaries.** For the most part we will be concerned with the location parameter problem with the following assumptions:

(2.1)

a) $X|\theta \sim f(x-\theta)$, where $\theta$ is the median of X, $f(\cdot)$ is a proper density with unbounded support, and has tails that are decreasing outside of a compact set,

b) $\theta \sim \pi(\theta)$, where the median of $\theta$, $\mu$, is taken to be zero without loss of generality. $\pi(\cdot)$ is a bounded density that gives positive mass to all points in its support.

Some of the following theorems will need additional assumptions on either f or $\pi$. These will be stated where needed.

We assume that $f(\cdot)$ has unbounded support mainly for convenience. Some of the theorems contained here can be made to apply to the bounded support case, but the type of reconciliation that we are looking for, as stated in (1.1), will not generally obtain.

Given the sampling density $f(x-\theta)$, the frequentist constructs a $1-\alpha$ confidence set $C(X) = \{\theta: |\theta-X| \le c\}$, where c is chosen to satisfy the frequentist probability constraint:

$$1-\alpha = P(\theta \in C(X)|\theta) = \int_{\{x:\theta\in C(x)\}} f(x-\theta)dx \quad . \tag{2.2}$$

Note that although we are taking C to be a symmetric set here, there is no loss in generality: any bounded set would serve as well. Using the prior $\pi(\theta)$, the Bayesian can make a post-data probability assessment of the confidence set C, its posterior credible probability, given by

$$P(\theta \in C(x)|x) = \frac{\int_{\{\theta:\theta\in C(x)\}} f(x-\theta)\pi(\theta)d\theta}{\int_{-\infty}^{\infty} f(x-\theta)\pi(\theta)d\theta} \quad . \tag{2.3}$$

Our concept of frequentist priors requires agreement between (2.2) and the limit of (2.3):

DEFINITION 2.1. For X and θ as in (2.1) and C(X) as in (2.2), we say that $\pi(\theta)$ is a *frequentist prior* with respect to C(X) if

$$\lim_{|x|\to\infty} P(\theta \in C(X)|X=x) = P(\theta \in C(X)|\theta) = 1-\alpha \quad .$$

Informally speaking, the prior must wash out as the data and the prior information become more discrepant. Thus, the specification of a frequentist prior is really only a specification of tail behavior of $\pi(\cdot)$ relative to $f(\cdot)$. We first can get some rather simple characterizations of frequentist priors.

THEOREM 2.1. Let X and θ be as in (2.1), and suppose that there exists a function g(X) such that

$$(2.4) \qquad \lim_{|x|\to\infty} \frac{\pi(x-y)}{g(x)} = 1 \quad \forall\, y \quad \text{and} \quad \int_{-c}^{c} \sup_{x} \frac{\pi(x-y)}{g(x)}\, dy < \infty \; .$$

Then $\pi$ is a frequentist prior for $C = \{\theta\colon |\theta-X| \le c\}$ if and only if

$$(2.5) \qquad \lim_{|x|\to\infty} \int_{-\infty}^{\infty} \frac{f(y)\pi(x-y)}{g(x)} dy = 1 = \int_{-\infty}^{\infty} \lim_{|x|\to\infty} \frac{f(y)\pi(x-y)}{g(x)} dy \quad .$$

PROOF. The posterior probability of C can be written as

$$P(\theta \in C(X)|X=x) = \frac{\int_{-c}^{c} f(y)\, \frac{\pi(x-y)}{g(x)} dy}{\int_{-\infty}^{\infty} f(y)\, \frac{\pi(x-y)}{g(x)} dy} \; .$$

As $|x| \to \infty$, the numerator clearly goes to $\int_{-c}^{c} f(y)dy = 1-\alpha$ by (2.4) and the Dominated Convergence Theorem. Therefore, the posterior coverage probability approaches one if and only if the denominator converges to one. ‖

Note that, in fact, $1-\alpha$ is an upper bound on the limiting posterior probability under the conditions of Theorem 2.1. This follows from Fatou's Lemma:

$$\liminf_{|x|\to\infty} \int_{-\infty}^{\infty} f(y) \frac{\pi(x-y)}{g(x)} dy \geq \int_{-\infty}^{\infty} \liminf_{|x|\to\infty} f(y) \frac{\pi(x-y)}{g(x)} dy = 1 \quad .$$

The existence of a function $g(\cdot)$ to satisfy conditions (2.4) appears to be a necessary condition for $\pi(\cdot)$ to be a frequentist prior, although the limit condition is minor, and is usually satisfied since c is finite. These conditions, however, are certainly not sufficient as the next theorem shows:

THEOREM 2.2. Let X and $\theta$ be as in (2.1). Let $g(\cdot)$ be any function satisfying (2.4). Suppose there exists a positive constant k such that

$$\lim_{x\to\infty} \frac{f(x+k)}{g(x)} > 0 \quad , \tag{2.6}$$

then $\pi$ is not a frequentist prior for f.

PROOF. From Theorem 2.1, if $\pi$ is a frequentist prior then (2.5) must be satisfied. This, in particular, implies that

$$\lim_{x\to\infty} \int_{x}^{x+k} f(y) \frac{\pi(x-y)}{g(x)} dy = 0 \quad . \tag{2.7}$$

However, for sufficiently large x,

$$\int_{x}^{x+k} f(y) \frac{\pi(x-y)}{g(x)} dy \geq \pi^{*}k \frac{f(x+k)}{g(x)} \quad , \tag{2.8}$$

where $\pi^{*} = \inf\{\pi(u): -k \leq u \leq 0\} > 0$ by (2.1b), and we have used the fact that f has decreasing tails outside a compact set. Using (2.8), the limit of the LHS of (2.7) is positive, contradicting (2.6) and proving the theorem. ‖

We can similarly prove an analog to Theorem 2.2, that refers to the lefthand tail of $f(\cdot)$. In particular, if $x \to -\infty$ and there exists a $k < 0$ satisfying (2.6), then $\pi(\theta)$ will not be a frequentist prior. (A similar comment applies to Theorem 2.3, below.)

We, therefore, see that the existence of a function $g(\cdot)$ satisfying (2.4) cannot be a sufficient condition for $\pi(\cdot)$ to be a frequentist prior. This was expected, however, since we would believe that a sufficient condition must somehow involve the sampling density $f(\cdot)$. Although we haven't tried to investigate all the possible limits of $\pi(x-y)/g(x)$, we have investigated one alternative to condition (2.4), and that alternative is given in the following theorem. Note that additional assumptions on both the prior and sampling density are required.

THEOREM 2.3. Let X and $\theta$ be as in (2.1), and, in addition, assume that f is symmetric and $\pi$ is unimodal and differentiable. If there exists a function $g(x)$ satisfying (2.4) and

a. $\lim_{x\to\infty} \frac{\partial}{\partial y} \frac{\pi(x-y)}{g(x)}$ converges uniformly for y in an open interval containing $[-c,c]$

b. $\lim_{x\to\infty} \frac{\pi(x-y)}{g(x)} = h(y)$, where $h(y) + h(-y)$ is nonconstant, $0 < y < c$,

then $\pi$ is not a frequentist prior with respect to $\{\theta : |\theta - X| \le c\}$.

The proof of Theorem 2.3 will rely on the following well-known lemma:

LEMMA 2.1. Let $f_1(y)$ and $f_2(y)$ be positive functions, and let $f_1(y)/f_2(y)$ be increasing and nonconstant for $0 < y < x$. Then

$$\frac{\int_0^x f_1(y)dy}{\int_0^\infty f_1(y)dy} < \frac{\int_0^x f_2(y)dy}{\int_0^\infty f_2(y)dy},$$

provided the integrals are finite.

PROOF of Theorem 2.3. As before, we can write the limiting posterior probability as

$$\lim_{x\to\infty} P(\theta \varepsilon C(X)|X=x) = \lim_{x\to\infty} \frac{\int_{-c}^{c} f(y)\frac{\pi(x-y)}{g(x)}dy}{\int_{-\infty}^{\infty} f(y)\frac{\pi(x-y)}{g(x)}dy}$$

$$\text{(2.9)} \qquad \le \frac{\int_{-c}^{c} f(y)h(y)dy}{\int_{-\infty}^{\infty} f(y)h(y)dy} ,$$

where we have used Dominated Convergence and (2.4) in the numerator and Fatou's Lemma in the denominator. By the symmetry of f, expression (2.9) is equal to

$$\text{(2.10)} \qquad \frac{\int_{0}^{c} f(y)[h(y)+h(-y)]dy}{\int_{0}^{\infty} f(y)[h(y)+h(-y)]dy} .$$

If the function h(y) + h(-y) were increasing, we could apply Lemma 2.1 with $f_1(y) = f(y)[h(y) + h(-y)]$ and $f_2(y) = f(y)$ to deduce that (2.10) is less than $1-\alpha$, and hence $\pi(\theta)$ is not a frequentist prior.

Thus, to complete the proof we need to show that h(y) + h(-y) is increasing, or equivalently,

$$\text{(2.11)} \qquad \frac{d}{dy}[h(y) + h(-y)] < 0, \qquad \text{for } |y| < c .$$

By the unimodality of $\pi$, $h(\cdot)$ is increasing and hence $h'(\cdot) > 0$. Therefore, (2.11) is automatically satisfied if $\frac{d}{dy} h(-y) = 0$. We now consider the case $\frac{d}{dy} h(-y) \neq 0$. Expression (2.11) is equivalent to

$$\text{(2.12)} \qquad \frac{h'(y)}{h'(-y)} > 1 \;\forall\, y, \qquad \text{for } |y| < c\,.$$

From the definition of $h(\cdot)$,

$$\text{(2.13)} \qquad \frac{h'(y)}{h'(-y)} = \frac{\frac{\partial}{\partial y} \lim\limits_{x\to\infty} [\pi(x-y)/g(x)]}{-\frac{\partial}{\partial y} \lim\limits_{x\to\infty} [\pi(x+y)g(x)]} = \frac{-\lim\limits_{x\to\infty}\left[\frac{\partial}{\partial y}\pi(x-y)/g(x)\right]}{\lim\limits_{x\to\infty}\left[\frac{\partial}{\partial y}\pi(x+y)/g(x)\right]}\,,$$

where this last interchange of limit and derivative follows from condition a (see, e.g., Theorem 13-13, Apostol, 1957, pg. 402). Since the last limit in (2.13) is equal to the limit of the ratio, we have

$$\begin{aligned} \frac{h'(y)}{h'(-y)} &= -\lim_{x\to\infty} \frac{\frac{\partial}{\partial y}\pi(x-y)}{\frac{\partial}{\partial y}\pi(x+y)} \\ \text{(2.14)} \qquad &= \lim_{x\to\infty} \frac{\frac{\partial}{\partial x}\pi(x-y)}{\frac{\partial}{\partial x}\pi(x+y)} \\ &= \lim_{x\to\infty} \frac{\pi(x-y)}{\pi(x+y)}\,, \end{aligned}$$

where the last equality follows from L'Hospital's rule. The unimodality of $\pi$ implies that $\pi(x-y) > \pi(x+y)$ for $x > y$, showing that the last limit in (2.14) is at least 1. Hence, (2.12) is established and the theorem is proved. ||

An application of Theorem 2.3 shows that the double exponential distribution can never be a frequentist prior. Indeed, if $X|\theta \sim n(\theta,1)$ and $\pi(\theta) = e^{-|\theta|/\beta}/(2\beta)$, then the limiting posterior probability of the set $C = \{\theta\colon |\theta-X| < c\}$ is

$$\lim_{x\to\infty} P(\theta \in C|x) = P\left(-c - \frac{1}{\beta} < Z < c + \frac{1}{\beta}\right) < 1-\alpha\quad,$$

where Z is a standard normal random variable.

Notice that condition (2.9) can be interpreted as saying that, for large x, we have

$$\pi(x-y) \approx g(x)h(y) \quad , \tag{2.15}$$

for some functions g and h. This is true for the double exponential prior, and no doubt is true for other priors as well. The important idea here is that if the prior "asymptotically factors" as in (2.15), it cannot be a frequentist prior.

With the possible exception of Theorem 2.3, the results in this section are quite broad. Specific cases tend to be somewhat more informative and have greater practical import. We look at some special cases in the next section.

**3. The normal and t-distributions.** In location parameter problems the most often assumed density (in classical statistics) is the normal density. Moreover, standard Bayesian textbook examples are often based on a normal-normal setup. Since the normal distribution has the sharpest tails among common (unbounded) densities, the fact that it fails as a frequentist prior should come as no surprise. This detail is taken care of in the following theorem.

THEOREM 3.1. Let $X \sim f(X-\theta)$ satisfy condition (2.1), and let $\pi(\theta) = n(0,1)$. The normal prior is not a frequentist prior for any $C = \{\theta: |\theta-X| \le c\}$ with positive confidence coefficient. In fact, $P(\theta \in C|x) \to 0$ as $|x| \to \infty$.

PROOF. An upper bound on $P(\theta \in C|x)$ can be obtained by

$$P(\theta \in C|x) = \frac{\int_{-c}^{c} f(y)\pi(x-y)dy}{\int_{-\infty}^{\infty} f(y)\pi(x-y)dy} \le \frac{\int_{-c}^{c} f(y)\left[\max_{|t|<c} \pi(x-t)\right]dy}{\int_{-\infty}^{\infty} f(y)\pi(x-y)dy}$$

$$= (1-\alpha) \Big/ \int_{-\infty}^{\infty} f(y) \frac{\pi(x-y)}{\max_{|t|\le c} \pi(x-t)} dy \quad . \tag{3.1}$$

As $|x| \to \infty$, if the limit of the denominator in (3.1) is greater than 1, then $\pi$ cannot be a frequentist prior. It suffices to consider only $x > 0$.

From Fatou's Lemma, $\lim_{x\to\infty} P(\theta \in C|x)$ is bounded above by (3.1). However, for the normal case

$$\lim_{x\to\infty} \frac{\pi(x-y)}{\max_{|t|\le c} \pi(x-t)} = \infty, \qquad \text{for } y > c \ ,$$

establishing the theorem. ||

A much more interesting special case is that of a t sampling density vs. a t prior density. Let $t_\nu(\cdot)$ denote the density of Student's t with $\nu$ degrees of freedom, and consider the sample density to be $f(x-\theta) = t_p(x-\theta)$ and the prior density to be $\pi(\theta) = t_q(\theta)$. Since the degrees of freedom controls the tail of the distribution (along with existence of moments), the relationship of the sample to prior degrees of freedom should determine when a particular prior is a frequentist prior. The following theorem completely specifies the situation.

THEOREM 3.2. Let $X-\theta \sim t_p(x-\theta)$ and $\theta \sim \pi(\theta) = t_q(\theta)$. Then $t_q$ is a frequentist prior if and only if $p > q$.

PROOF. For the function $g(x) = t_q(x)$ we have

$$\lim_{|x|\to\infty} \frac{\pi(x-t)}{g(x)} = \lim_{|x|\to\infty} \frac{t_q(x-t)}{t_q(x)} = 1 \quad ,$$

so, by Theorem 2.1, it is sufficient to examine the behavior of

$$\lim_{x\to\infty} \int_{-\infty}^{\infty} t_p(y) \frac{t_q(x-y)}{t_q(x)} \, dy \quad . \tag{3.3}$$

First consider $p \leq q$. It is easy to check that for any positive constant k

$$\lim_{x\to\infty} \frac{t_p(x+k)}{t_q(x)} > 0 \qquad \text{if} \qquad p \leq q \quad , \tag{3.4}$$

so an application of Theorem 2.2 will show that $t_q$ cannot be a frequentist prior if $p \leq q$. (In fact, if $p < q$ then the limit in (3.4) is infinite, showing that $P(\theta \in C|x) \to 0$ as $x \to \infty$. For $p > q$, we need to show that the limit in (3.3) is equal to 1. Fix $\varepsilon$, $0 < \varepsilon < 1$, and divide the region of integration into two parts: $\{y: |y| < \varepsilon|x|\}$ and $\{y: |y| > \varepsilon|x|\}$. We now establish that, if $p > q$, then

$$(3.5) \qquad \lim_{|x|\to\infty} \int_{\{y:|y|>\varepsilon|x|\}} t_p(y) \frac{t_q(x-y)}{t_q(x)} dy = 0$$

and

$$(3.6) \qquad \lim_{|x|\to\infty} \int_{\{y:|y|<\varepsilon|x|\}} t_p(y) \frac{t_q(x-y)}{t_q(x)} dy = 1 \quad .$$

Combining (3.5) and (3.6) will show that (3.3) is equal to 1, establishing that $t_q$ is a frequentist prior.

To establish (3.5), note that the integral is trivially bounded by

$$\frac{t_p(\varepsilon|x|)}{t_q(x)} \int_{-\infty}^{\infty} t_q(x-y)dy = \frac{t_p(\varepsilon|x|)}{t_q(x)} \quad ,$$

which goes to zero as $|x| \to \infty$. To establish (3.6), first note that for $|y| < \varepsilon|x|$, $t_q(x-y) < t_q[(1-\varepsilon)|x|]$ and also

$$\lim_{x\to\infty} \int_{-\infty}^{\infty} t_p(y) \frac{t_q[(1-\varepsilon)|x|]}{t_q(x)} I_{(-\varepsilon|x|,\varepsilon|x|)}(y)dy$$

$$= \int_{-\infty}^{\infty} \lim_{x\to\infty} t_p(y) \frac{t_q[(1-\varepsilon)|x|]}{t_q(x)} I_{(-\varepsilon|x|,\varepsilon|x|)}(y)dy \quad ,$$

by directly calculating both sides. This allows us to apply a generalized Dominated Convergence Theorem (Billingsley, 1986, Exercise 16.6) to pull the limit inside the integral in (3.6) and establish the equality. ||

The results of Theorem 3.2 have implications in case of estimating the mean of a normal distribution with unknown variance. In particular, if $X_1,\cdots,X_n$ are iid $n(\theta,\sigma^2)$, consider a prior distribution

$$\pi(\theta,\sigma^2) = t_q(\theta) \frac{1}{\sigma} d\theta d\sigma \quad ,$$

where we take a noninformative prior on $\sigma$. Since $\sigma$ is really a nuisance parameter here, a noninformative prior seems justified. A standard calculation will show that the distribution of $(\bar{X},S^2)$ given $\theta$ (unconditional on $\sigma$), is proportional to

$$\left(1 + \frac{n(\bar{x}-\theta)^2}{(n-1)s^2}\right)^{-n/2} ,$$

which behaves like a t with n-1 degrees of freedom. If we now consider $s^2$ fixed, then we are in the case of Theorem 3.2, and any prior on $\theta$ that is flatter than a $t_{n-2}$ will be a frequentist prior. (Note that this is not the standard frequentist inference, which would be unconditional on $\bar{X}$ and $S^2$.)

**4. Comments and Generalizations.** The idea of a frequentist prior is not just an oxymoron, but rather a legitimate means for a frequentist to make post-data probability statements while retaining a frequency flavor in the analysis. Use of a frequentist prior necessitates subjective input, but requires that the subjective input never be weighted more than the data. Furthermore, use of the frequentist prior is more satisfying than the use of a noninformative prior. With the latter type of prior, post-data probability statements are made with respect to an arbitrary distribution, one which may have no meaning whatsoever in the context of a given problem.

If one starts with a $1-\alpha$ confidence set $C(X)$, then the post-data probability, with respect to a frequentist prior, may be greater than $1-\alpha$ for some data values, but is necessarily less than $1-\alpha$ for some data values (since the priors are proper). This fluctuation is illustrated in Figure 1, which shows post-data probabilities for a t-sampling density using different t priors. Control of the amount of variation from the nominal $1-\alpha$ has not been addressed here: this is a "second-order" phenomenon and will be the basis for future work.

There are other questions which can be raised, and we will only comment on two of them. One is a question of unimodality of the posterior, and how it relates to the flatness of the prior. In general, we feel that experimenters are most comfortable with unimodal posterior densities, or densities that are essentially unimodal. Such posteriors seem to occur when the tails of the sample density and prior are different. We have no theorems concerning unimodality, but offer Figure 2, which shows the situation for two configurations of t densities. In the two cases shown, the posterior distributions are not unimodal, showing that whether or not the tails of the sample and prior distributions match, the posterior need not be unimodal. In contrast, Figure 3 shows unimodal posteriors arising from a normal sample and double exponential prior distributions with very different tails.

A second, more practical question, has to do with multiple observations. For the most part we have dealt with one observation from a location density. If more than one observation is collected, the distribution of the sufficient statistic may not be expressible as a location distribution. (If the observations come from a location family that admits a unique maximum likelihood estimator, this estimator has a location distribution and our results apply to confidence sets centered around it.) In general, however, the question of existence of frequentist priors for distributions other than those in the location family must be explored.

As was noted in the text, the results presented here cover sampling from a normal distribution, so the multiple observation case is taken care of in this situation. In fact, if one considers sampling from a multivariate normal density, the results on frequentist priors carry over to this case also, if we interpret $|X-\theta|$ as a multi-dimensional Euclidean norm.

Another simple generalization is to confidence sets of forms other than $\{\theta: |X-\theta| \le c\}$. Our results hold for any confidence set that collapses to this one as $|X-\theta| \to \infty$. In particular, the Stein-type confidence sets considered by Casella and Hwang (1987) have the same frequentist priors as the usual confidence set.

## REFERENCES


APOSTOL, T. M. (1957). *Mathematical Analysis: A Modern Approach to Calculus*. Addison-Wesley, Reading, MA.

BILLINGSLEY, P. (1986). *Probability and Measure*, Second Edition. John Wiley and Sons, NY.

BUEHLER, R. J. (1959). Some validity criteria for statistical inference, *Ann. Math. Statist.* **30** 845-863.

CASELLA, G. (1985). Conditionally acceptable recentered set estimators, Technical Report No. BU-889-M, Biometrics Unit, Cornell University, Ithaca, NY 14853.

CASELLA, G. (1986). Conditionally acceptable frequentist solutions, Technical Report No. BU-908-M, Biometrics Unit, Cornell University, Ithaca, NY 14853. To appear in the *Proceedings of the Fourth Purdue Symposium on Decision Theory* (with discussion).

CASELLA, G. and HWANG, J. T. (1987). Employing vague prior information in the construction of confidence sets. *Journal of Multivariate Analysis* **21** 79-104.

DAWID, A. P. (1973). Posterior expectations for large observations. *Biometrika* **60**, 664-666.

FISHER, R. A. (1956). *Statistical Methods and Scientific Inference*, Third Edition. Hafner, New York.

KIEFER, J. C. (1977). Conditional confidence statements and confidence estimators (with discussion), *J. Amer. Statist. Assoc.* **72** 789-827.

LeCAM, L. (1984). Discussion of "The Likelihood Principle" in *The Likelihood Principle: A Review, Generalizations, and Statistical Implications* by J. Berger and R. Wolpert. Monograph Series, Institute of Math. Statist.

MEEDEN, G. and ISAACSON, D. (1977). Approximate behavior of the posterior distribution for a large observations. *The Annals of Statistics* **5**, 899-908.

PIERCE, D. A. (1973). On some difficulties with a frequency theory of inference, *Ann. Statist.* **1** 241-250.

ROBINSON, G. K. (1979). Conditional properties of statistical procedures, *Ann. Statist.* **7** 742-755.

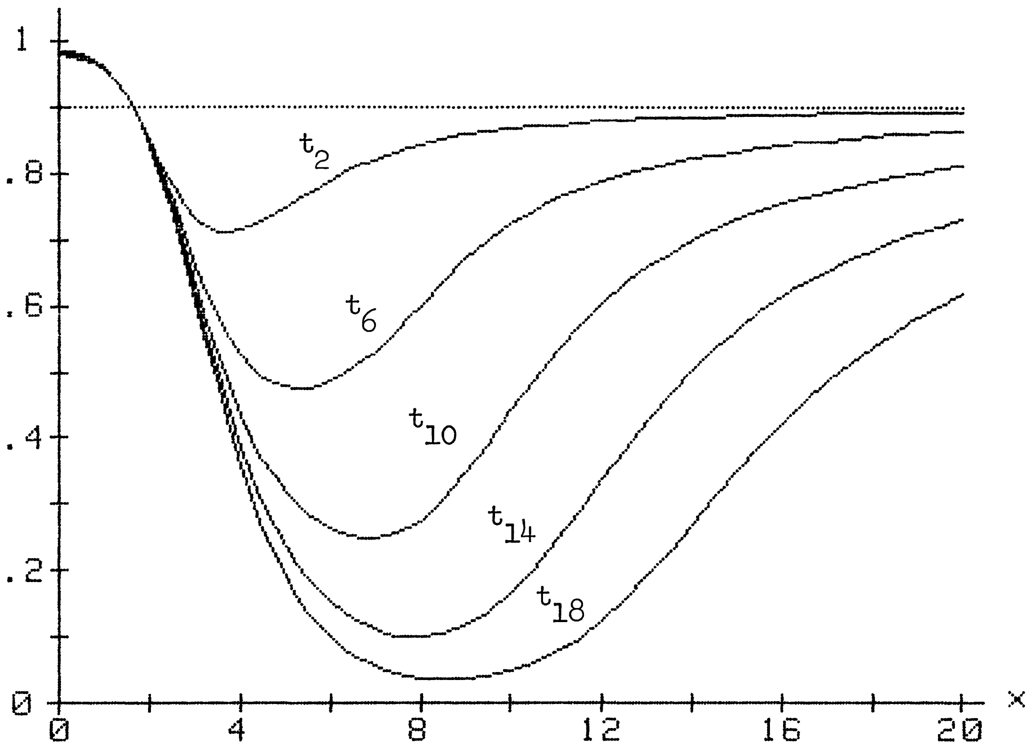


FIG. 1. Posterior probabilities for $t_{10}$ sampling density and five different t prior densities.

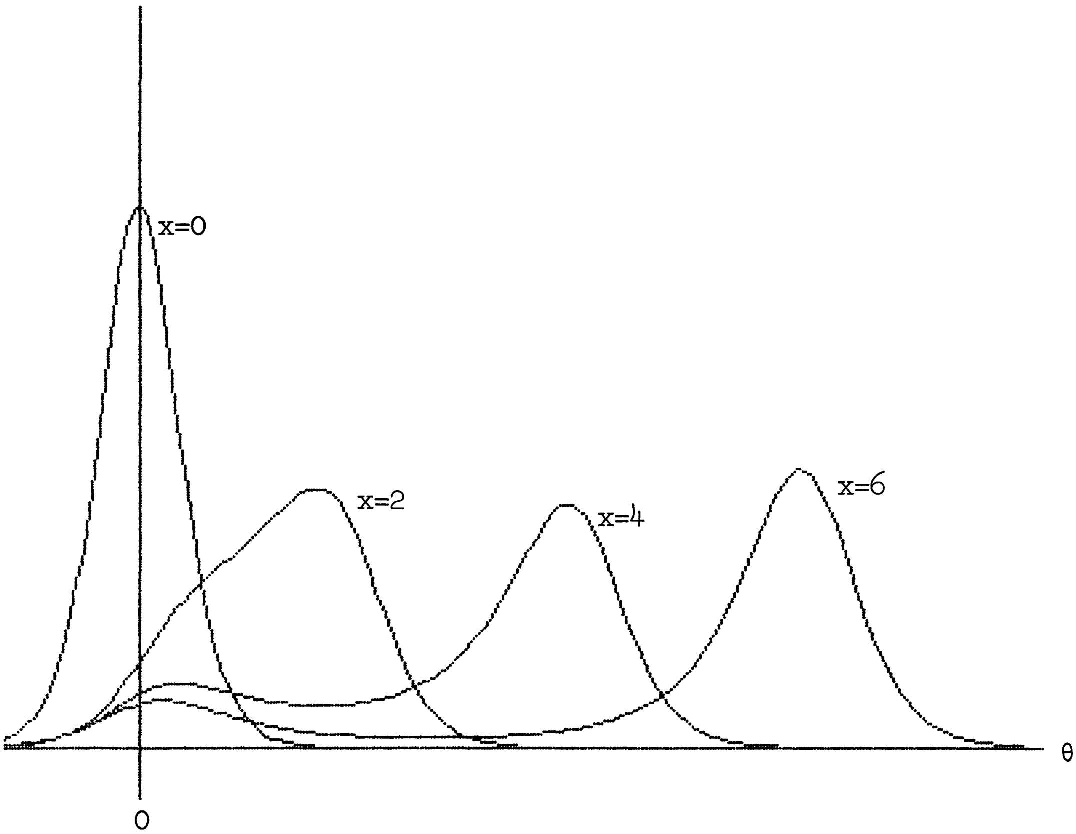


FIG. 2a. Posterior densities from $t_4$ sampling density and $t_3$ (mean zero) prior density.

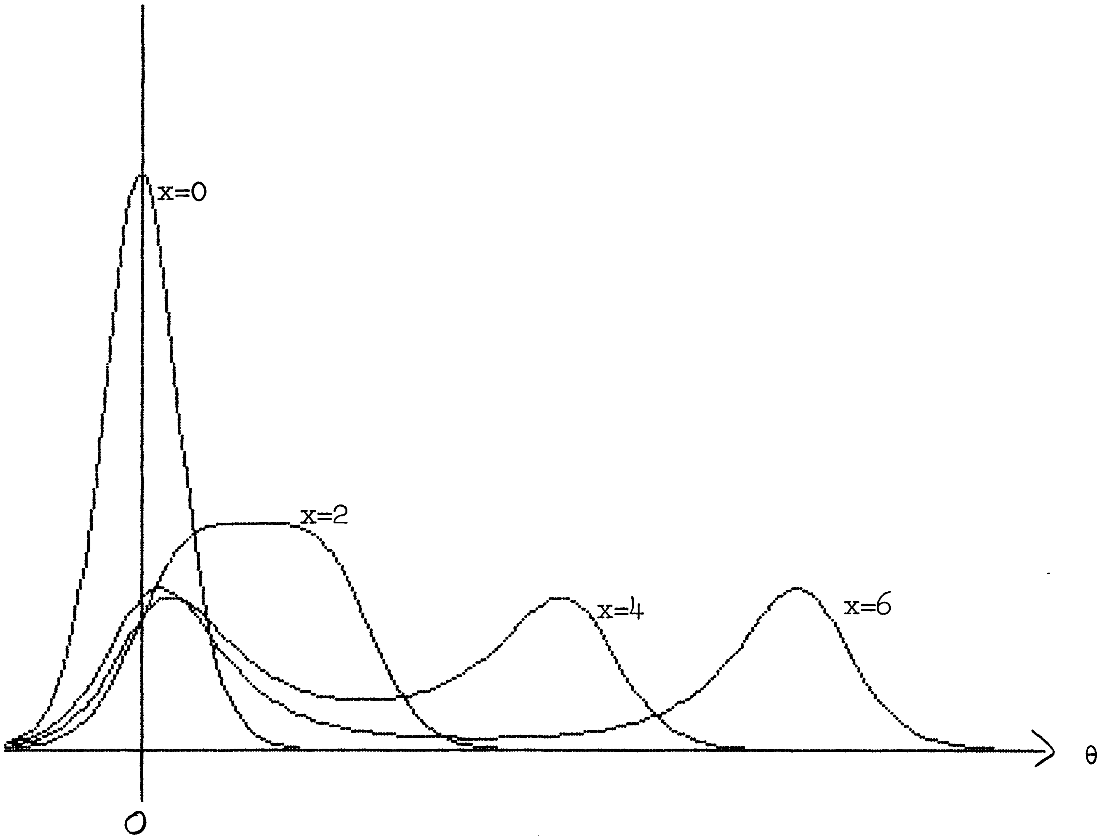


FIG. 2b. Posterior densities from $t_4$ sampling density and $t_4$ (mean zero) prior density.

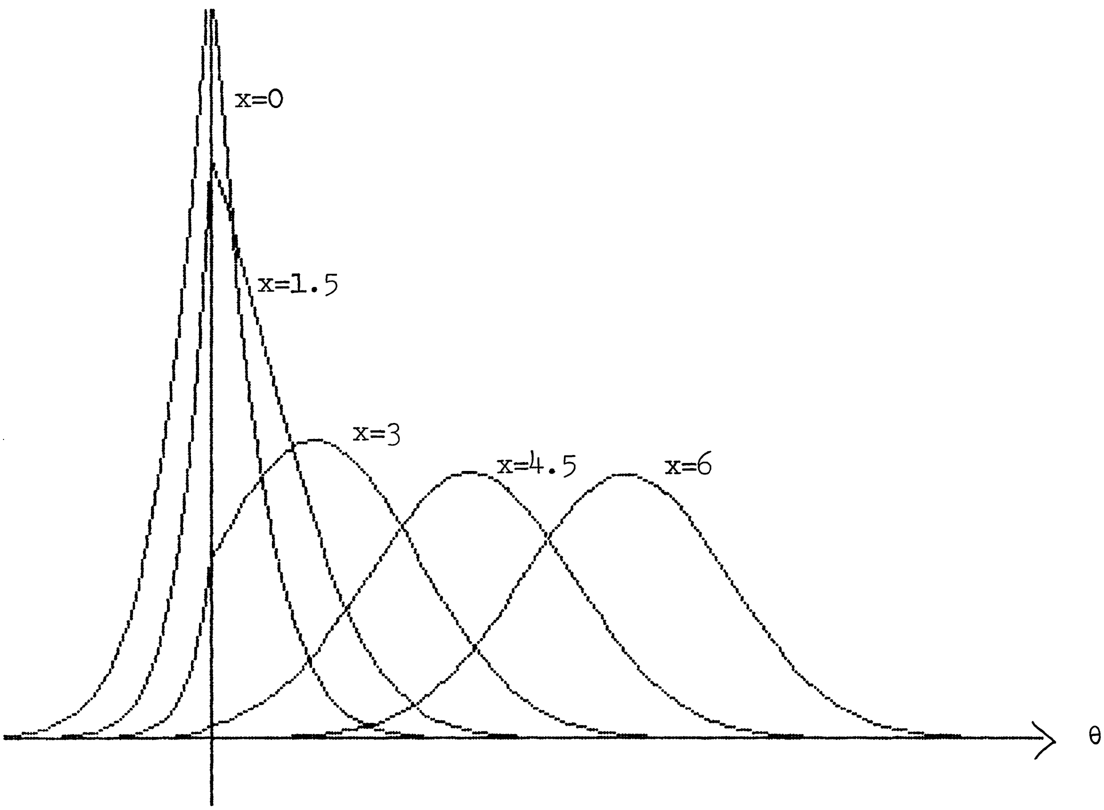


FIG. 3. Posterior densities from normal sample density and double exponential (mean zero) prior density.